A Dynamical Behavior Measurement Algorithm for Smart Meter Data: An Analytical Study

Abbas M. Al-Ghaili
Institute of Informatics and Computing in Energy *(IICE)*
Universiti Tenaga Nasional *(UNITEN)*
Kajang, Malaysia
abbas@uniten.edu.my

Hasventhran Baskaran
Institute of Informatics and Computing in Energy *(IICE)*
Universiti Tenaga Nasional *(UNITEN)*
Kajang, Malaysia
hasventhran@gmail.com

Zul-Azri Ibrahim
Institute of Informatics and Computing in Energy *(IICE)* & College of Computing &Informatics *(CCI)*
Universiti Tenaga Nasional *(UNITEN)*
Kajang, Malaysia
zulazri@uniten.edu.my

Fiza Abdul Rahim
Institute of Informatics and Computing in Energy *(IICE)* & College of Computing &Informatics *(CCI)*
Universiti Tenaga Nasional *(UNITEN)*
Kajang, Malaysia
fiza@uniten.edu.my

Syazwani Arissa Shah Hairi
College of Computing &Informatics *(CCI)*
Universiti Tenaga Nasional *(UNITEN)*
Kajang, Malaysia
ak95128@ut.edu.my

***Abstract*— Smart Meter Data (SMD) has been considered by many as sensitive due to many reasons. One of these reasons is the security issue possessed by the SMD. With the raising popularity of the smart meter deployment around the world have raised new issues regarding the security of the SMD. Attacker with malicious intention and equip with certain capability can directly inject a false reading to the actual consumption of the SMD. Because of this reason, the protection and privacy of SMD have become a consideration in many researches. In order to protect and secure SMD, it is very important to understand the pattern of the SMD behavior. The studied from SMD can expose the past event that happened to the actual data as any direct manipulation to the SMD. In order to track any changes to the SMD, this paper has proposed a Dynamical Behavior Measurement Algorithm (DBMA), which involves a number of mathematical operations in order to understand and highlight in graphs the behavior of SMD and the mechanism of numbers' changing. The obtained results from DBMA would be useful to track the SMD if there is any form of malicious data being injected to the actual SMD thus can help in enhancing the SMD security. An experimental work has been conducted where 43200 samples have been tested and analysed using the proposed DBMA. The analysis evaluation of this work has been presented in this paper.**

Keywords*— *Dynamical behavior, Smart meter data, Smart meter data analytics, Time series

I. INTRODUCTION

Smart meters have the capabilities to automatically send the consumer reading information to the utilities provider using an open network [1]. The SMD included consumer voltage, current, phase and the power factor in each of specific time. Not like traditional analogue meter, the smart meter data can be used in various type of purposed [2-10]. The data can give important information about the load profile and the power consumption behavior other related considerations [11-13] for consumer. With all the data information contain in the SMD, can give a lot of benefit to the utilities providers in determining the effective energy tariff structure and the demand response operation [14] for each of the consumer.

Technically the way smart meter communicate between each other is by digitally sending information from one meter to another hop by hop. Because of the way the data move is the same as computer base communication network, smart meter also possess the vulnerabilities as normal computer does. According to [15] not only terrorist that is interested in inflicting damage to smart grid environment, the beginner hacker/cracker is lured to experiment their skill in launching attacks to smart meter environment and one of the popular attacks is the False Data Injection Attack (FDIA). Authors in [16] have mentioned that if an attacker failed to breach the communication line between smart meter, they will be able to create their own smart meter and physically install the spoof meter to transmit a false value to manipulate the SMD.

This paper is organized as follows: the proposed DBMA is in detail presented in Section III. Analysis is explained in Section IV. In Section V, Discussion has been provided. Conclusion is drawn in Section VI.

II. RELATED WORKS

There is a number of research work [17, 18] that had introduced techniques to be used as defend against FDIA. Authors in [17] have introduced a statistical method for detecting FDIA by applying the Kullback–Leibler Distance (KLD) that analyzed the historical data. The research work in [18] has proposed an approach to utilize the correlation and measurement history with the sparse attribute of malicious attack to detect the FDIAs. Other methods have been used to protect their original data and smart meters reading consumption by tracing the dynamic behavior of measured data so that it can verify if original data that has been tampered in an unauthorized manner.

There are different methods proposed to perform a dynamical behavior procedure aiming to understand the mechanism of numbers that had changed. For example, an approach by using clustering of electricity consumption behavior dynamic toward the big data application is presented in [19]. In this technique the Fast Search and Find of Density Peaks (FSFDP) is integrated into a divide and conquer approach toward big data. Time-based Markov model is implemented to formulate the dynamics of consumer electricity consumption behavior in both centralized and distributed clustering process.

Another example was introduced by [20] in which the author proposed to deduce the user behavior from SMD due to user activities in a household. The author used the Non-Intrusive Load Monitoring (NILM) algorithm to break down the total of electrical energy used based on SMD reading and the data recovery mechanism will take place in case of poor data quality. Then, the NILM algorithm was applied to deduce the used of appliances from the recovered data. A natural language generator tool has been developed to convert the process data into textual report and presenting it as the final result.

The SMD clustering using consumption indicator introduced by [21] whereas, in this study, authors have outlined the SMD application to identify the consumer with peak system. The consumption variance was used as an indicator of behavior variability by implementing two clustering techniques, Hierarchical clustering and Self-organizing map (SOM). Both approaches is to evaluate the characteristics of the customer group and to leverage the SOM to visualization the cluster.

The authors in [22] introduced robust SMD Analysis using Smoothed Altering Least Squares (SALS) and Dynamic Time Warping (DTW) approach. The SALS approach will recover corrupted and missing data by learning spatial-temporal properties of data and improve the accuracy while the DTW approach will cluster large-scale daily load curves with different sampling rate. Thus, using both approach the SMD are warped non-linearly in the time dimension to determine a measure of similarity independent of certain non-linear variations in the time dimension.

Application and analysis of SMD along with RTL SDR and GNU Radio was proposed by [23]. The author focused on improving the performance of the system by integrating Realtek Software Defined Radio (RTLSDR) and GNU radio along with the smart meter. The efficiency of the setup is enhanced by monitoring the high frequency Electromagnetic Interference (EMI) produced by the electrical appliances. EMI signal will neglect the surrounding noise generated along with SMD to analyze the mode of operation of each electrical appliance.

In [24], the SMD analytics for optimal customer selection in demand response (DR) program uses the real data. DR program, "Time of Use (ToU)" help to improve generation capacity when load demand increase. The SMD easily detected when DR events are declared. The data are used by the DR program may lead user to modify their own consumption data to low peak-hour consumption and user increase their off-peak demand.

In regard to time series consideration, some methods have used the Euclidean distance to analyze SMD for electricity consumers and to solve spike corresponding to switching on/off [25]. Some other researches have applied AMI smart meter big data analytics for time series of electricity consumption [26]. It presents analysis of 5-minutes SMD sets to explore time series of electricity consumption and create forecast models, Auto Regressive Integrated Moving Average (ARIMA), Exponential Smoothing which resulted in a lower prediction error.

W. Luan [27] when doing smart meter data analytics for distribution network connectivity verification have mentioned that a lot of utilities has data quality problem with the geographical information system (GIS) records at distribution level. To correct this connectivity errors in the GIS, BC Hydro has developed an in-house algorithm where it can leverages smart meter interval measurements and identifies the meters by voltage profile correlation analysis for various scenario. The first three scenarios has showcase the application of the algorithm. The fourth scenario showcases the results from another urban feeder. The fifth scenario analyses the potential of using the same algorithm for identifying feeder phases of customer meters in a downstream section of a rural feeder. The algorithm performs well when applied to customers attached to single phase transformers via over-head connections.

One of the limitations is the dynamic behavior measuring and reading procedure. This is useful because it helps to understand the way that numbers change. When the change in values are obvious, it would be helpful to track the dynamical behavior of smart meter numbers flow. This can aid in the process to detect injected numbers and values. This paper has proposed Dynamical Behavior Measurement Algorithm (DBMA) that has its own dynamical behavior measuring way which might be affected by several conditions. The proposed algorithm consists of several steps implemented in a sequential order. For every step, extracted features will mainly play important role in the way its dynamical behavior evolves. When such a designed algorithm is applied on other measure data which different features, performance may fail to achieve accurate results. Therefore, this paper aims to design a simple algorithm that tracks the dynamical behavior of measured data obtained from real scenarios.

III. The Proposed Dynamical Behavior Measurement Algorithm (DBMA)

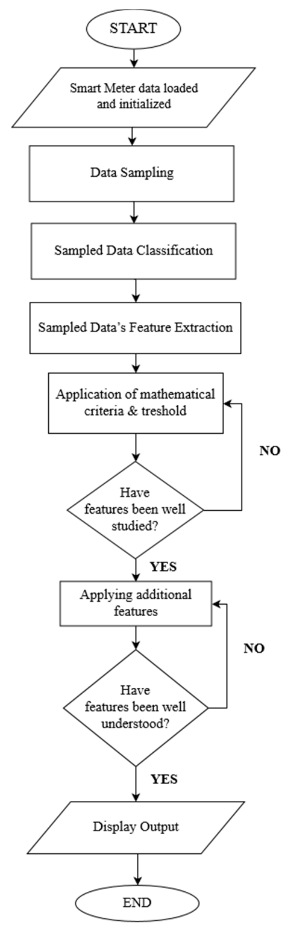

Figure 1. The proposed flowchart of DBMA

A. Overview

Half-hourly basis datasets for electricity consumption of London based smart meter were chosen to be analysed in this study. These datasets were provided by CACI International Inc's consumer classification software, Acorn. Many blocks with multiple smart meters' half-hourly usage data were found in the datasets. A block was selected randomly from the data blocks. Out of the 45 smart meters in the chosen block, 10 smart meters were chosen randomly. Data sampling and analysis were performed on these 10 smart meter's half-hourly usage datasets. The proposed flowchart is illustrated in Figure 1 above.

B. Proposed Mathematical Procedure for Total Number of Datasets and Samples

Each smart meter has 1440 samples each of which is a measured energy usage value that is read from the selected smart meter. Thus, a single smart meter represents whole reads for a month. That is, number of samples for each smart meter during the period, p_d of a month (i.e., 30 days) can be calculated using Eq. (1):

$$n_i = p_d \times h \times f \tag{1}$$

Where,

- n is the total number of samples for i^{th} smart meter
- i represents no. of smart meter; $i = 1$ to 10
- p_d is the total period in days; $p_d = 30$
- h is the total number of hours; $h = 24$
- f is the basis factor on which every SM_i measures the energy value for each hour. Meaning, how many times, at each hour (h), reads will be taken for each smart meter is recorded by f; $f = 2$ in this research study.

Therefore,

$$n_1 = 30 \times 24 \times 2$$
$$= 1440 \text{ samples}$$

To calculate whole no. of samples which have been used for experiments, Eq. (2) is used:

$$n_T = T_{SM} \times p_m \times n_{sm_i} \tag{2}$$

where,

- n_T is the total number of sample for all smart meters
- T_{SM} is the total number of smart meters
- p_m is the total period in months from which samples have been taken
- n_{sm_i} is the total number of samples for each smart meter, as derived from Eq. (1). Thus, Eq. (2) is applied to obtain the total number of samples for all smart meters, as follows:

$$n_T = 10 \times 3 \times 1440$$
$$= 43{,}200 \text{ samples}$$

To calculate the total number of samples, that is derived from Eq. (2), as formulated in Eq. (3):

$$S_T = T_{SM} \times p_m \tag{3}$$

where n_T represents the total number of datasets. Thus, there are $10 \times 3 = 30$ datasets comprising 43,200 samples.

C. The proposed DBMA Flowchart

Ten London based Smart Meter data was chosen to measure the dynamical behavior of datasets provided by smart meters. The data was sampled for each Smart Meter for three months. The data of a Smart Meter for each month was kept as a distinctive dataset to differentiate it with other two months. The data for each month was classified on half-hourly basis. Hence, there are 1440 data entry for each month from one Smart Meter. Based on the existing data, summation and average was performed to statistically understand the data behavior/ consumption pattern for each month. A mathematical criterion was applied to find thresholding values such as minimum and maximum. If the features were well studied, the flow continues to the next step. If there is any lack of understanding or correctness in the feature, the flow reverses back to step applying mathematical criteria. Additional features can be applied in the next step. Relation between designed parameters and features is understood and relation between summation and square of max-min energy values can be found through Eq. (4):

$$V_{SM} = \frac{n \times \left(s + \frac{s}{n}\right)}{\frac{1}{2}(max \times min)} \times \sqrt{\frac{max \times min}{s}} \tag{4}$$

where,

- V_{SM} is an obtained featured value that defines amount of total number of samples multiplied by the summation in relation to maximum and minimum obtained within each dataset
- n is the total number of samples for a dataset
- s is the summation function for whole samples within one dataset
- max is maximum value obtained within the selected dataset
- min is minimum value obtained within the selected dataset.

The obtained value, Eq. (4), is a feature extracted in order to highlight a distinguished value derived from main parts of measured SMD. That is, the V_SM defines mathematically the amount of no. of samples associated with its related summation in regard to maximum and minimum values. This computed value is obtained once Eq. (4) has been applied to each dataset. Thus, each dataset will produce a unique value that can distinguish the dataset. Values obtained from all datasets, in our experiment, are similar but not identical. Therefore, once the dataset has been attacked, Eq. (4) will produce a different value which can be discoverable.

IV. Analysis

This analysis has used five factors in order to evaluate the performance of the proposed DBMA.

A. Average Value Factor

As can be seen in Figure 2, it is noticeable that, for most of samples and measures of smart meters, averaged values of energy range between 1×10^{-1} and 3×10^{-1}.

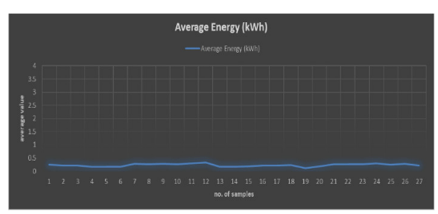

Figure 2. Average measured energy values for datasets

B. *The Sum Factor*

The sum values for each sample can be, for whole 30 datasets, between 200 and 400. Meaning, $200 \leq sum \leq 400$. In Figure 3, all related values are provided.

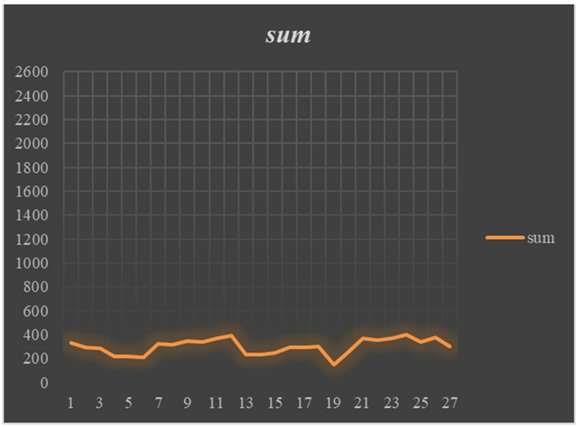

Figure 3. *sum* values for measured energy for whole samples per each dataset; 30 datasets

C. *Relation between the sum factor and n*

It is clear that the relation between *sum* and *n* is an inverse proportion. In Figure 4, *n* values for whole datasets are mentioned. At each y-axis, the value represents *no. of samples*.

Figure 4. n-values for measured energy for 30 datasets

D. *Relation between sum and max and min energy value for each smart meter*

In this type of analysis, a mathematical operation-based equation has been designed in order to understand the way the numbers dynamically behave and change. This equation creates a variable in order to compute the rate between sum values for each smart meter for each month with a relation to maximum and minimum values. It is mathematically formulated in Eq. (5):

$$V_{SM} = \frac{\sqrt{2s}}{\sqrt{max \times min}} \times (n+1) \quad (5)$$

This equation is similar to Eq. (4) and has been derived from Eq. (4). It has a less computation time. Once this equation has been applied on every monthly-basis value for each smart meter, results are shown in Fig. 5. It is found that values are located between 100,000 and 300,000. The V_{SM} is to have a rate between *sum* energy values to half of *max* and *min* values with a relation to a square root of *max* and *min* values. It is designed to add a special feature to our datasets and samples being tested. Its aim to allow us to understand the dynamical behavior of energy values (kWh), so that values can be traced and evaluated for a potential analysis.

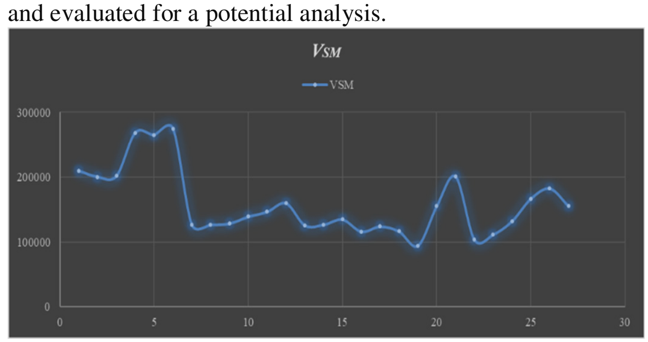

Figure 5. Rate values for the V_{SM}

E. *Tracking factor based on* $\frac{max \times min}{2}$ *to the* $\boldsymbol{V_{SM}}$ *factor*

By tracing the path of behavior of obtained values of the formula $\frac{max \times min}{2}$ in regard to the V_{SM} mentioned in Figure 5, it is found that the relation between both of them is an inverse proportion.

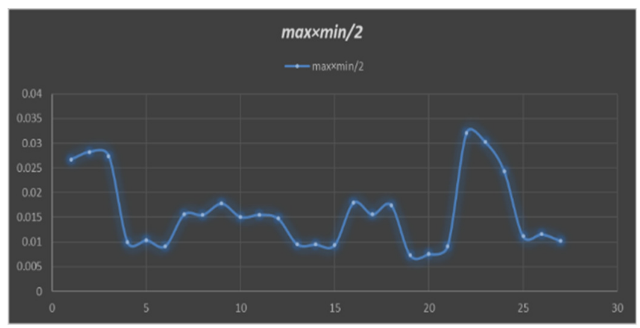

Figure 6. The dynamical behavior of $\frac{(max \times min)}{2}$ values related to V_{SM}

By comparing Figure 5 to Figure 6, it is seen that when the slope between two values in Figure 5 is negative, its corresponding slope value on Figure 6 is a positive value. As for example, by focusing on values on points 22, 23, 24, 25, and 26 shown in Figure 5, they are gradually increased; and in same time they decrease in Figure 6.

V. Discussion

This proposed study has designed a mathematical procedure to track the dynamical behavior of SMD reads and the way their related reads (energy measured values and reads in kWh). As can be seen in Fig. 2 and Fig. 5, the dynamical behavior locates under a known range of numbers. The proposed equations which are Eq. (4) and Eq. (5), have been implemented on whole samples and the obtained results are shown in Fig. 5 and Fig. 6. The proposed procedure mentioned by Eq. (4) and Eq. (5) will produce different values of energy (kWh) once original data has been in an unauthorized manner modified. Therefore, the privacy of SMD (43,200 samples) would be affected. The proposed DBMA is a useful tool to help detect SMD's change.

VI. CONCLUSION

The dynamical behavior of Smart Meter Data (SMD) has been studied. In order to understand how SMD's measured values dynamically behave and how energy reads and numbers change, there is a need to track numbers individually and then measure the relation between numbers. For this purpose, a DBMA has been proposed to perform a measurement purpose. The proposed DBMA has applied a series of math-based operations to measure and derive relation(s) between each SMD's number with other numbers. The relation has been graphically represented to show the way the SMD behaves.

The obtained results have shown that, for tested datasets, the relation between average of (max and min) is reversely proportional with the combination of sum and half of max and min. this is an important tracking-based monitoring procedure to protect SMD form such a falsified data injection occurrence. The evaluated and analyzed results showed that the proposed DBMA is suitable for SMD and it might contribute to detect such an action attempting to change smart meter data. Thus, the proposed DBMA could be exploited by some other SMD protection techniques.

VII. ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Higher Education Malaysia (MoHE) and Universiti Tenaga Nasional (UNITEN) for funding this study under the Fundamental Research Grant Scheme (FRGS) of Grant No. FRGS/1/2017/ICT03/UNITEN/03/1

REFERENCES

[1] V. V. Das, "Wireless Communication System for Energy Meter Reading," in 2009 International Conference on Advances in Recent Technologies in Communication and Computing, 2009, pp. 896-898.

[2] S. Wan Yen, S. Morris, M. A. G. Ezra, and T. Jun Huat, "Effect of smart meter data collection frequency in an early detection of shorter-duration voltage anomalies in smart grids," International Journal of Electrical Power & Energy Systems, vol. 109, pp. 1-8, 2019/07/01/ 2019.

[3] A. Brint, G. Poursharif, M. Black, and M. Marshall, "Using grouped smart meter data in phase identification," Computers & Operations Research, vol. 96, pp. 213-222, 2018/08/01/ 2018.

[4] R. Razavi, A. Gharipour, M. Fleury, and I. J. Akpan, "Occupancy detection of residential buildings using smart meter data: A large-scale study," Energy and Buildings, vol. 183, pp. 195-208, 2019/01/15/ 2019.

[5] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster Analysis of Smart Metering Data," Business & Information Systems Engineering, vol. 4, pp. 31-39, February 01 2012.

[6] M. E. H. Dyson, S. D. Borgeson, M. D. Tabone, and D. S. Callaway, "Using smart meter data to estimate demand response potential, with application to solar energy integration," Energy Policy, vol. 73, pp. 607-619, 2014/10/01/ 2014.

[7] B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, "Recent advances in the analysis of residential electricity consumption and applications of smart meter data," Applied Energy, vol. 208, pp. 402-427, 2017/12/15/ 2017.

[8] A. Al-Wakeel, J. Wu, and N. Jenkins, "k-means based load estimation of domestic smart meter measurements," Applied Energy, vol. 194, pp. 333-342, 2017/05/15/ 2017.

[9] K. Hopf, M. Sodenkamp, I. Kozlovkiy, and T. Staake, "Feature extraction and filtering for household classification based on smart electricity meter data," Computer Science - Research and Development, vol. 31, pp. 141-148, August 01 2016.

[10] L. Wen, K. Zhou, S. Yang, and L. Li, "Compression of smart meter big data: A survey," Renewable and Sustainable Energy Reviews, vol. 91, pp. 59-69, 2018/08/01/ 2018.

[11] U. C. C. Jagstaidt, J. Kossahl, and L. M. Kolbe, "Smart Metering Information Management," Business & Information Systems Engineering, vol. 3, p. 323, August 20 2011.

[12] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in smart grids: a review," Energy Informatics, vol. 1, p. 8, 2018/08/13 2018.

[13] N. Uribe-Pérez, L. Hernández, D. De la Vega, and I. Angulo, "State of the Art and Trends Review of Smart Metering in Electricity Grids," Applied Sciences, vol. 6, p. 68, 2016.

[14] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," IEEE Transactions on Power Systems, vol. 19, pp. 1232-1239, 2004.

[15] J. C. Foreman and D. Gurugubelli, "Cyber Attack Surface Analysis of Advanced Metering Infrastructure," arXiv preprint arXiv:1607.04811, 2016.

[16] R. Mahmud, R. Vallakati, A. Mukherjee, P. Ranganathan, and A. Nejadpak, "A survey on smart grid metering infrastructures: Threats and solutions," in 2015 IEEE International Conference on Electro/Information Technology (EIT), 2015, pp. 386-391.

[17] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in ac state estimation," IEEE Transactions on Smart Grid, vol. 6, pp. 2476-2483, 2015.

[18] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting False Data Injection Attacks on Power Grid by Sparse Optimization," IEEE Transactions on Smart Grid, vol. 5, pp. 612-621, 2014.

[19] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications," IEEE Transactions on Smart Grid, vol. 7, pp. 2437-2447, 2016.

[20] E. Ebeid, R. Heick, and R. H. Jacobsen, "Deducing Energy Consumer Behavior from Smart Meter Data," Future Internet, vol. 9, p. 29, 2017.

[21] M. Azaza and F. Wallin, "Smart meter data clustering using consumption indicators: responsibility factor and consumption variability," Energy Procedia, vol. 142, pp. 2236-2242, 2017/12/01/ 2017.

[22] Z. Jiang, D. Shi, X. Guo, G. Xu, L. Yu, and C. Jing, "Robust Smart Meter Data Analytics Using Smoothed ALS and Dynamic Time Warping," Energies, vol. 11, p. 1401, 2018.

[23] M. Aswathi, R. Gandhiraj, and K. P. Soman, "Application and Analysis of Smart Meter Data along with RTL SDR and GNU Radio," Procedia Technology, vol. 21, pp. 317-325, 2015/01/01/ 2015.

[24] M. Martinez-Pabon, T. Eveleigh, and B. Tanju, "Smart Meter Data Analytics for Optimal Customer Selection in Demand Response Programs," Energy Procedia, vol. 107, pp. 49-59, 2017/02/01/ 2017.

[25] G. Dudek, A. Gawlak, M. Kornatka, and J. Szkutnik, "Analysis of Smart Meter Data for Electricity Consumers," in 2018 15th International Conference on the European Energy Market (EEM), 2018, pp. 1-5.

[26] R. Mohammad, "AMI Smart Meter Big Data Analytics for Time Series of Electricity Consumption," in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2018, pp. 1771-1776.

[27] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart Meter Data Analytics for Distribution Network Connectivity Verification," IEEE Transactions on Smart Grid, vol. 6, pp. 1964-1971, 2015.